



Lewis, E., Quinn, N., Blenkinsop, S., Fowler, H. J., Freer, J., Tanguy, M., Hitt, O., Coxon, G., Bates, P., & Woods, R. (2018). A rule based quality control method for hourly rainfall data and a 1km resolution gridded hourly rainfall dataset for Great Britain: CEH-GEAR1hr. *Journal of Hydrology*, 564, 930-943.
<https://doi.org/10.1016/j.jhydrol.2018.07.034>

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
[10.1016/j.jhydrol.2018.07.034](https://doi.org/10.1016/j.jhydrol.2018.07.034)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via Elsevier at <https://doi.org/10.1016/j.jhydrol.2018.07.034> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>



Research papers

A rule based quality control method for hourly rainfall data and a 1 km resolution gridded hourly rainfall dataset for Great Britain: CEH-GEAR1hr



Elizabeth Lewis^{a,*}, Niall Quinn^b, Stephen Blenkinsop^a, Hayley J. Fowler^a, Jim Freer^{b,c}, Maliko Tanguy^d, Olivia Hitt^d, Gemma Coxon^b, Paul Bates^{b,c}, Ross Woods^{c,e}

^a School of Engineering, Newcastle University, Newcastle Upon Tyne NE1 7RU, UK

^b School of Geographical Sciences, University of Bristol, Bristol BS8 1SS, UK

^c Cabot Institute, University of Bristol, Bristol BS8 1UJ, UK

^d Centre for Ecology & Hydrology, Wallingford OX10 8BB, UK

^e Department of Civil Engineering, University of Bristol, Bristol BS8 1TR, UK

ARTICLE INFO

This manuscript was handled by Marco Borgia, Editor-in-Chief, with the assistance of Francesco Marra, Associate Editor

Keywords:

Gridded hourly precipitation

Extremes

Quality control

Great Britain

Dataset

Rain gauge

ABSTRACT

High-resolution gridded precipitation products are rare globally, particularly below a daily time-step, yet many hydrological applications require, or can be improved by, a higher temporal resolution of rainfall data. Here, we present a new 1 km resolution gridded hourly rainfall dataset for Great Britain (Gridded estimates of hourly areal rainfall for Great Britain (1990–2014) [CEH-GEAR1hr]) using data from over 1900 quality controlled rainfall gauges, which improves upon the current UK national gridded precipitation datasets at daily time-step. We extend and automate a quality control (QC) procedure to permit the use of hourly data for 1990–2014 and independently validate the QC using daily rainfall data and recorded historic events. Our two-tiered validation approach, at daily and hourly timescales, indicates that spurious extreme values are excluded from the resultant dataset, while legitimate values are preserved. We use a nearest neighbour interpolation scheme to derive gridded hourly rainfall values at 1 km resolution, to temporally disaggregate the CEH-GEAR daily gridded dataset and produce an hourly dataset with consistent daily totals. This provides a unique resource for hydrological applications in Great Britain. The CEH-GEAR1hr dataset, associated metadata and QC information, will be freely available from the Environmental Information Data Centre (EIDC) and hosted alongside the daily and monthly CEH-GEAR product.

1. Introduction

Recently, increased attention has been given to sub-daily precipitation observations due to the contribution of intense rainfall events to flash flooding in urban areas and fast-responding catchments. Indeed, our ability to address and plan for flash floods has been partly limited by the paucity of available high-quality rain gauge data (Westra et al., 2014). Several studies have demonstrated the sensitivity and improved performance of hydrological model simulations when driven by precipitation data at a sub-daily time step (e.g. Finnerty et al., 1997; Bastola and Misra, 2013), particularly for small catchments with rapid response times. Additionally, the lack of temporal resolution offered by daily data for direct application in flood forecasting and the need for assessment of the impacts of short-duration intense rainfall events on

hydrological systems has created a requirement for the improved availability of sub-daily precipitation data. This was further identified as a need in the World Climate Research Programme Grand Challenge on Extremes (Alexander et al., 2016) and the INTENSE project has taken up the mantle of collecting and quality-controlling a global sub-daily precipitation dataset (www.research.ncl.ac.uk/intense). Such datasets are also invaluable for the validation of the new generation of very high-resolution convection-permitting climate models (CPMs – see Prein et al. (2015) for a review) which offer improved representation of sub-daily extreme rainfall (e.g. Kendon et al., 2012; Kendon et al., 2014; Chan et al., 2014 for the UK).

Gridded datasets, providing spatially comprehensive information on weather and climate variables, have wide-ranging applications. These include providing the inputs to national hydrological modelling (e.g.

Abbreviations: CEH-GEAR1hr, Gridded estimates of hourly areal rainfall for Great Britain (1990–2014); CEH-GEAR, Gridded estimates of areal rainfall; QC, Quality Control

* Corresponding author.

E-mail address: Elizabeth.lewis2@ncl.ac.uk (E. Lewis).

<https://doi.org/10.1016/j.jhydrol.2018.07.034>

Received 23 March 2018; Received in revised form 9 June 2018; Accepted 14 July 2018

Available online 20 July 2018

0022-1694/ © 2018 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Bell et al., 2007; Yang et al., 2014), the assessment of historical climate and its variability (e.g. Blenkinsop et al., 2008; Simpson and Jones, 2014; Becker et al., 2013; Yu et al., 2016) and the assessment of reanalysis and downscaled climate model products (e.g. Isotta et al., 2015). Existing datasets of ground-based observations are typically on daily timescales, but gridded hourly products offer the potential for enhanced applications in these areas as well as for the verification of quantitative precipitation forecasts, satellite products and the validation of CPMs. Radar data offers the required temporal resolution but suffers from errors in the estimation of precipitation magnitude (Collier, 1989; Villarini and Krajewski, 2010) and may lead to reduced performance of hydrological models compared with gauge-derived data (Cole and Moore, 2009; Parkes et al., 2013).

A number of gridded precipitation products derived from rain gauges are available for the UK and are summarised in Table S1 in the Supplementary Information (comprising UK-only and Europe-wide datasets). Reanalysis data, which are created by a data assimilation scheme and models which incorporate observations at 6–12 h time-steps, may also be used to characterise the observed long-term variability of precipitation (e.g. NCEP/NCAR: Kalnay et al., 1996; 20CR: Compo et al., 2011; ERA-Interim: Dee et al., 2011). However, these are typically at coarser spatial resolutions and are not appropriate for most hydrological modelling studies. For example, Rhodes et al. (2015) noted that although reanalysis products represent many of the features of large-scale precipitation and daily totals over England and Wales, individual extreme events are less well represented. Regional reanalyses may address some of the problems associated with coarse scale reanalysis datasets though improved understanding of uncertainties is needed (Borsche et al., 2015).

Gridded datasets at hourly resolution have been constructed for some regions using a range of methodological approaches although these are typically only for short time-periods, and often constructed for the calibration and assessment of hydrological models in catchments. For example, Wüest et al. (2010) used a dense daily rain-gauge network for Switzerland for 1992–2003 and disaggregated this to the hourly timescale using radar data to preserve daily totals. A similar approach was used by Paulat et al. (2008) to create a gridded hourly dataset for Germany for the period 2001–2004. Shen et al. (2014) merged over 30,000 hourly gauge observations with satellite data to produce a gridded hourly dataset for China for the 2008–2010 warm seasons. An alternative approach was applied by Luo et al. (2013) who constructed an hourly precipitation grid over the Yangtze–Huai Rivers Basin in China for the 2007 “mei-yu” season using a direct interpolation method with automated weather station (AWS) data. Some national scale sub-daily datasets have been produced however. Vormoor and Skaugen (2013) developed a 1 km 3 h gridded precipitation dataset for Norway using a model simulated hindcast series to disaggregate an existing gridded daily dataset for the period 1957–2010. Hourly observations for approximately 2,500 stations in the US have also been interpolated onto a relatively coarse (2° latitude by 2.5° longitude) grid (Higgins et al., 1996) and a gridded product created through merging of daily gauges with radar (Cosgrove et al., 2003) up to a resolution of $1/8^\circ$.

For the UK, generation of gridded sub-daily precipitation data on a catchment-scale over limited periods has been performed on an ad hoc basis for the assessment of hydrological models and their input data. Typically, this has been achieved by disaggregating a relatively dense daily rain gauge network using radar (e.g. Parkes et al. (2013) for the 4062 km² of the Upper Severn River; Cole and Moore (2008) for 136 km² of the River Darwen and 212 km² of the River Kent catchments). The availability of quality-controlled individual hourly rain gauge data for the UK (Blenkinsop et al., 2017, and extended here) offers the potential for a more extensive gridded hourly dataset for UK coverage.

Quality control of gauge rainfall data is essential to ensure a high quality product. There are many shortcomings of gauge measurement of rainfall including mechanical errors, recording errors, evaporation

from partly-filled buckets, wind-induced under-catch, and snow-effects (McMillan et al., 2012). QC procedures can identify some of these errors, for example, frequent tips (Upton and Rahimi, 2003), erroneous extreme values and erroneous dry periods (Abbott, 1986). However, QC procedures are unable to identify errors due to undercatch and evaporation. Therefore the measurements may still not reflect the true amount of rain that fell even after QC checks are completed. Quality control is typically a manual operation and in some instances quasi-automated, but still relies heavily on either a final manual inspection or on previously manually identified errors for model training. Prior to this work, analysis of the UK rain gauge data has focussed only on a subset (< 20%) of available rain gauges due to the need for near-complete records for climatological analysis and the labour-intensive nature of quality controlling such data (Blenkinsop et al., 2017).

This paper builds on Blenkinsop et al. (2017) by: i) developing additional quality control procedures whose implementation may be automated and applied to all ~1900 gauge records, and ii) using these data to produce a gridded (1 km) hourly dataset for the UK for 1990–2014. In Section 2 we describe the data sources, Section 3 describes the extended and automated quality control procedure for the hourly rain gauge data, which includes a rule base for the implementation of single-site and nearest neighbour gauge checks. Section 4 details the validation of the automated QC process. Section 5 describes the resulting gauge dataset. Section 6 describes the methodological basis for the disaggregation of the daily gridded dataset. Section 7 assesses the reliability of the hourly gridded dataset and finally, in Section 8 we discuss the implications and potential applications of this dataset and potential future avenues which could improve the product and be used to investigate the associated uncertainties.

2. Data

2.1. CEH-GEAR 1 km daily dataset

To generate gridded hourly precipitation values we adopt a similar approach to that used in other studies (as discussed in Section 1), by disaggregating an existing quality controlled, validated gridded daily precipitation dataset. Here we use the CEH Gridded Estimates of Areal Rainfall (CEH - GEAR) dataset (Keller et al., 2015; Tanguy et al., 2016), which is an open source dataset that provides 1 km gridded estimates of daily and monthly rainfall for Great Britain and Northern Ireland from 1890 to 2014. The rainfall estimates in this dataset are derived from UK Met Office rain gauge observations, which were gridded using a natural neighbour interpolation method (a smooth, weighted version of nearest neighbour interpolation). We extracted the daily gridded estimates for 1990–2014 and temporally disaggregated these using an expanded version of the quality-controlled hourly rain gauge dataset of Blenkinsop et al. (2017). Thus, the new, hourly gridded dataset, CEH-GEAR1hr, preserves the daily totals of, and is consistent with, the widely used CEH-GEAR dataset. This allows the hourly gridded dataset to be made open source (currently the hourly rain gauge data is not freely available) and also allows direct comparisons to be made between the different temporal resolutions.

2.2. Hourly rain gauge data

To disaggregate the daily gridded dataset, a large dataset of hourly accumulations from 1903 UK rain gauges (see Fig. 1 for coverage) was used including a mixture of tipping bucket (TBR), 15 min and hourly data. These data were derived from the UK Met Office Integrated Data Archive System (MIDAS – downloadable from the British Atmospheric Data Centre, Met Office, 2012), the England Environment Agency (EA), Natural Resources Wales (NRW) and the Scottish Environmental Protection Agency (SEPA). This dataset was originally created and used in Blenkinsop et al. (2017). For use in this paper, the dataset was extended to the end of 2014 and an additional 216 gauges were included (all of

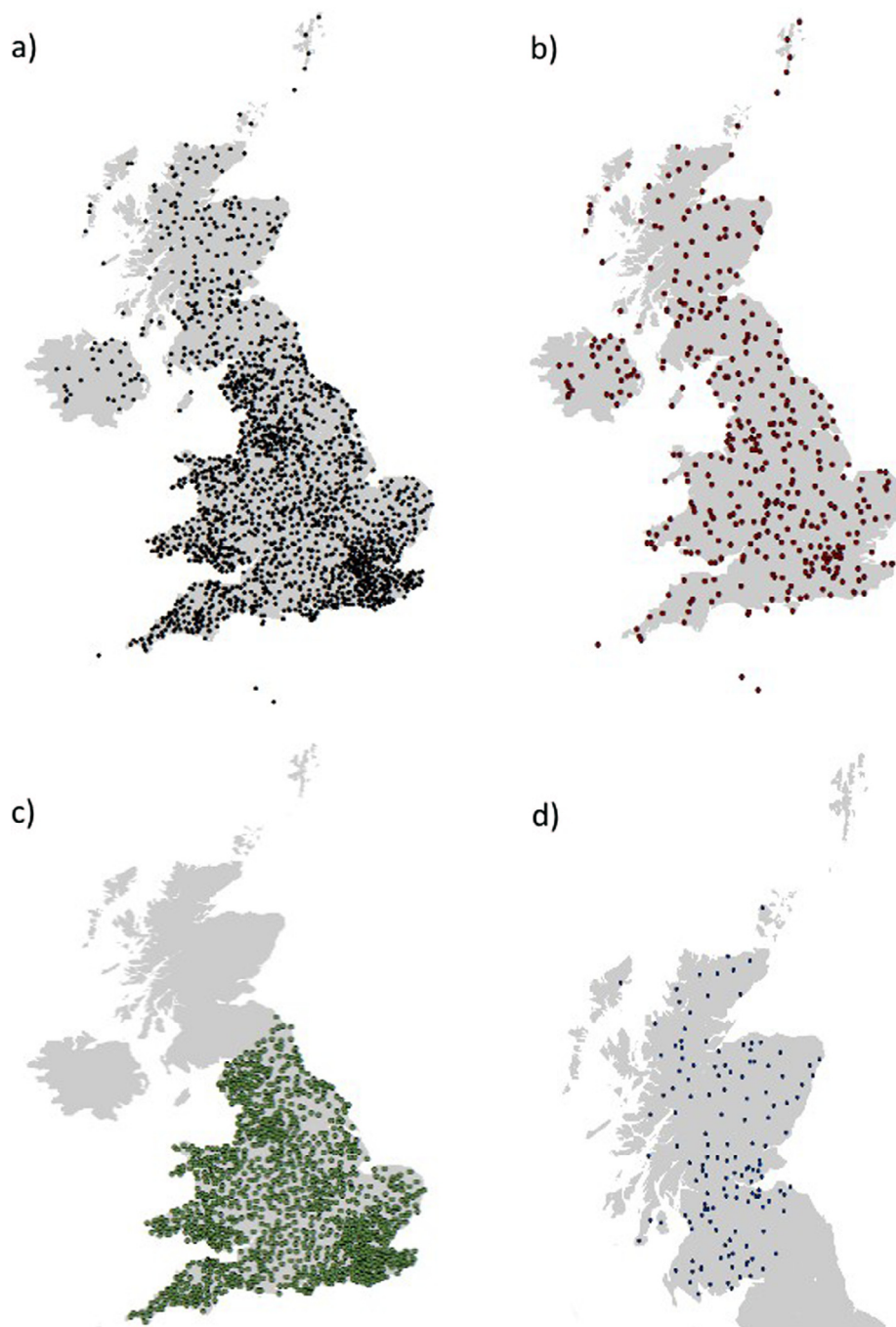


Fig. 1. Rain gauges providing hourly rainfall data for the UK during the period 1990–2014. Data is shown for a) all data sources, b) MIDAS, c) EA and NRW and d) SEPA.

less than 10 years duration) from MIDAS.

Blenkinsop et al. (2017) identified significant errors within these data and noted that the original dataset required additional quality control (QC) due to problems which included the recording of accumulated totals (principally daily), unfeasibly large hourly and daily values, unrecorded non-operation of gauges and, in the case of TBR data, unrealistic, high frequency tipping. In particular, internal QC procedures employed by the EA, including the use of check gauges, were noted to be affected by the institutional administrative structure and also varied in time. A series of single-site gauge tests were applied to identify the most egregious errors; however, a significant amount of time-consuming manual inspection of gauge records was required to produce the Blenkinsop et al. (2017) dataset, which comprises a subset

of 376 near-complete quality controlled gauges covering the period 1992–2011. These limited QC procedures identified two clear requirements to enable the maximum value to be derived from the entire dataset. Firstly, the incorporation of additional QC methods to include those which use neighbouring gauge data as an additional check. These are routinely used for the QC of daily data (e.g. Keller et al., 2015; Sciuto et al., 2009) but their application on hourly timescales is problematic given the localised nature of convective storms. Secondly, the development of an automated QC ‘rule base’ that integrates the results of the various QC tests, to negate the requirement for large-scale manual data inspection. This requires the identification of appropriate metrics by which to assess the performance of the rule base and the overall quality of the dataset, and these are described in Section 4. This

also provides the potential to further develop these formalised rules to either improve QC procedures as more understanding of the data is realised, to investigate the uncertainty associated with the QC process, or to have a different set of rules for specific applications.

3. Automated quality control tests

The quality control procedure is a three step process:

- 1) The gauge data is compared to the gridded CEH-GEAR daily dataset to identify *suspect gauges* which may subsequently be excluded from the dataset.
- 2) A series of quality control tests are applied to identify *suspect values* at all gauges which are marked with a quality control flag (but not excluded at this stage).
- 3) Combinations of quality control flags for a given hourly accumulation (referred to here as ‘rules’) are used to determine which flagged data are treated as *erroneous values* and thus excluded from the gauge records.

3.1. Step 1. Comparison of gauge data to gridded daily dataset

We first compared the (pre-QC) hourly gauge data (accumulated to 24 h) to the CEH-GEAR gridded daily dataset to estimate the initial quality of the rain gauge data and provide a baseline against which the QC process may be assessed. For each gauge we selected the CEH-GEAR grid cell over its location and used the Spearman’s rank correlation coefficient (r_s) and percentage correct statistics (P11, P00) (Wilks, 2006, Yoo and Ha, 2007) for comparison of the two time series. The latter is here calculated as the proportion of days in which rainfall is greater than 0 mm in both records (P11) or is dry in both records (P00) and is therefore a measure of concordance in rainfall occurrence. Whilst CEH-GEAR may not provide an accurate representation of rainfall at a point location, as it is an interpolated product and may also contain its own errors, the comparison is helpful to highlight potential errors in the gauge records which can then be examined further (recorded as suspect by QC9 in Table 2). Fig. 2 demonstrates that the majority of gauges (92.9%) match well with the corresponding grid cell (where $r_s > 0.8$ and $P00 + P11 > 0.8$). However, the long tail of gauges outside these bounds highlights the need for rigorous quality control of the gauge data.

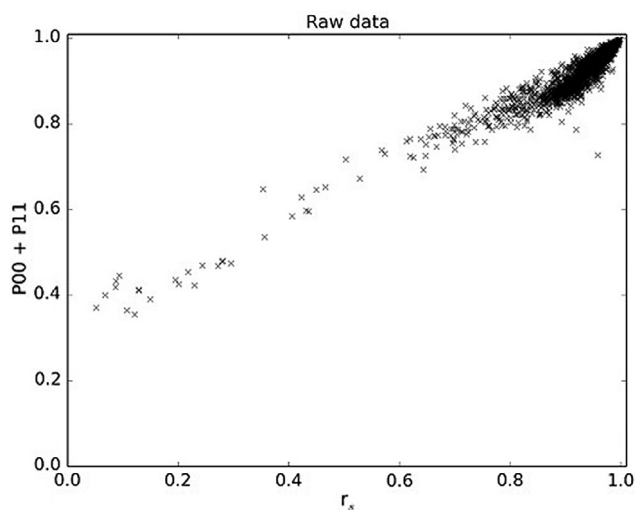


Fig. 2. Comparison of hourly gauge data (accumulated to 24 h) with the corresponding CEH-GEAR daily time series for 1903 gauges. $P00 + P11$ is the proportion of days in which it rains in both records or it is dry in both records, r_s is the Spearman's rank correlation coefficient.

3.2. Step 2. Identification and flagging of suspect gauge data

Blenkinsop et al. (2017) applied a series of single-gauge QC tests independently to each rain gauge. We supplement this by comparing the hourly data with that of neighbouring gauges. In total therefore, we apply 15 QC tests: 11 single-gauge QC tests and 4 neighbourhood-gauge QC tests (identifying dry spells and high values, each applied seasonally), with each hourly value allocated a flag if a potentially suspect hourly/daily total is identified.

The single-site gauge QC tests are based on the understanding of rainfall processes and known measurement practice (see Blenkinsop et al., 2017 for further details). For example, we include checks against known rainfall records and checks for common instrumentation errors such as accumulations (Table 1, flags 1 to 11).

Some types of errors, such as reporting or instrument errors, are unlikely to be duplicated across gauge networks and are more readily detected by comparison with neighbouring gauges. We use neighbourhood analysis to assess whether measurements at a gauge of interest are statistical outliers when compared with those of their similar neighbours (Table 1, QC flags 12 to 15). Such approaches are most useful when the correlation decay distance is high, e.g. for temperature, but have also been used previously in the QC of rainfall datasets, see e.g. Eischeid et al. (1995), Upton and Rahimi (2003), Sciuto et al. (2009) and Keller et al. (2015). Given the high variability in hourly rainfall totals, we apply these techniques to 24 h (daily) aggregations. See Supplementary Information for a detailed description of the neighbourhood analysis methodology.

3.3. Step 3. Application of a rule base to exclude suspect values

One of the key objectives of this work was to develop a quasi-automated procedure to interpret data flagged as potentially suspect in order to minimise the need for manual intervention in the QC process. This would mean that the QC procedure could be a) modified relatively easily and efficiently, b) the number of rules applied could be changed for different situations or analyses if less strict criteria were required, and c) the QC automated process could be applied to other comparable rain gauge datasets. The rule base presented in Table 2 uses the QC flags determined in the previous step (Table 1) in an intelligent manner based on knowledge of regional rainfall processes and characteristics and common errors in the rain gauge data (Blenkinsop et al., 2017). It thus comprises a set of 20 rules (R_n) that combine the QC flags either individually, in combination with other flags or in relation to other data characteristics such as dry sequences. Its aim is to fulfil the two criteria of excluding the most egregious errors in the data but also simultaneously preserving ‘real’ extreme values. Whilst the QC flags applied in step 2 provide valuable information on suspect data in the dataset, individual flags may not identify erroneous data effectively. For example, the threshold-based tests (QC1-3) are derived from the UK Met Office gauge network but this does not include the additional EA gauge data that significantly increases the coverage of the UK and may capture previously unrecorded events. A judgement is therefore required for such tests as to the appropriate thresholds at which a value may confidently be judged to be erroneous. The use of different thresholds allows the testing of different levels of severity in the implementation of the QC tests. Here, we judge that marginal threshold exceedance is insufficient on its own to identify erroneous data and so only automatically exclude such data if the UK record is exceeded by at least 20% (Table 2, R_1). For smaller exceedances, data is only excluded if further evidence of problems with the data exists (Table 2, $R_2 - R_7$) as is the case for those values with non-threshold flags ($R_8 - R_{13}$).

The application of the neighbouring gauge checks also needed careful consideration. Initially these were implemented throughout the year but it was noted that this resulted in the exclusion of a well-documented event at Boscastle, south west England in 2004 (Doe, 2004). As noted previously, the most intense hourly rainfall typically

Table 1

Summary of automated quality control flags applied to all data and described in the main text (QC process step 2).

Acronym	Description	Flag applied
Quality control flags applied to suspect hourly precipitation values		
QC1 (Threshold)	1 h record ¹ (92 mm) with separate flags showing exceedance by:	
	< 20%,	1
	≥ 20%,	2
	33%,	3
	or 50%	4
QC2 (Threshold)	Intense 'winter' rainfall with 1 h total > 80% of 1 h record (92 mm) ¹ in October–April period.	1
QC3 (Threshold)	24 h record ¹ (279 mm) with separate flags showing exceedance by:	
	< 20%,	1
	≥ 20%,	2
	33%,	3
	or 50%	4
QC4 (Non-threshold)	Suspect daily accumulations at 0900 or 1200 flagged where a recorded rainfall amount at these times is preceded by 23 h with no rain. A threshold of 2 × the mean wet day amount for the corresponding month is applied to increase the chance of identifying accumulated values at the expense of genuine, moderate events:	
	Accumulation at 0900	1
	Accumulation at 1200	2
QC5 (Non-threshold)	Suspect consecutive daily accumulations at 0900 or 1200 flagged recorded rainfall amounts at these times are preceded by 23 h with no rain on consecutive days with no threshold to the wet hour amount applied.	
	Accumulation at 0900	1
	Accumulation at 1200	2
QC6 (Non-threshold)	Suspect monthly accumulations. Identified where only one hourly value is reported over a period of a month and that value exceeds the mean wet hour amount for the corresponding month (a lower threshold than in QC4 is used here as a dry month is much more unlikely than a dry day in GB).	1
QC7 (Non-threshold)	Total consecutive 'large' values. Large values are defined as those exceeding 2 × the mean wet hour amount for the corresponding month. This threshold was applied as genuine duplicate small to moderate rainfall amounts occur legitimately.	1
QC8 (Non-threshold)	Frequent tipping (only applied to EA TBR data using Upton and Rahimi (2003) algorithm).	1
QC9 (Non-threshold)	Additional suspect data identified by manual inspection and comparison with documentary records e.g. British Rainfall or the CEH-GEAR daily data.	1
Flags applied to dry hours		
QC10	'Terminal' dry spell at start/end of gauge record ≥ 31d duration	1
QC11	Dry period:	
	≥ 31 d duration,	1
	≥ 45 d duration	2
Flags from neighbourhood analysis		
QC12	Neighbourhood check – dry periods > 14 days. Winter (NDJFMA):	
	90th,	1
	95th,	2
	99th bounds	3
QC13	Neighbourhood check – dry periods > 14 days. Summer (MJJASO):	
	90th,	1
	95th,	2
	99th bounds	3
QC14 (Threshold)	Neighbourhood check – daily precipitation total. Winter (NDJFMA):	
	90th,	1
	95th,	2
	99th bounds	3
QC15 (Threshold)	Neighbourhood check – daily precipitation total. Summer (MJJASO):	
	90th,	1
	95th,	2
	99th bounds	3

¹ Here the record used is the national record for the relevant timescale (1 h or 24 h) as acknowledged by the UK Met Office (refer to <http://www.metoffice.gov.uk/public/weather/climate-extremes> for a list of UK climatological extremes).

occurs from late spring to early autumn (Blenkinsop et al., 2017) and can be highly localised. This causes a rapid decrease in correlation between gauges with distance in summer months resulting in: i) daily accumulations differing significantly to the surrounding neighbours, ii) dry spells that are not recorded in surrounding neighbours. To allow for this highly localised nature of extreme rainfall events in summer months, the implemented rule base only applies high threshold value neighbourhood checks in winter (R20).

The final hourly precipitation gauge dataset includes the provision of all QC flags as gauge metadata. This ensures that all QC decisions are both transparent and traceable, and that users are able to test alternative rules and apply a custom rule base if required as appropriate for their application. For example, if gauge data were to be assimilated with radar data, the occurrence of rainfall may be useful even if the magnitude is erroneously recorded.

4. Validation of automated QC process

The QC rule base aims to exclude erroneous data whilst retaining correct values. Definitive validation of the rule base is impossible to achieve as other sources of national rainfall data may also contain errors. However, some attempt at validation is essential to provide confidence in the resultant dataset. We have therefore validated the rule base in two ways, firstly by comparing the resultant quality-controlled gauge data to a different gridded daily dataset, and secondly by comparing it to known historic storm and flood events recorded in the literature.

4.1. Validation of the exclusion of large rainfall values

Daily totals were calculated from the quality-controlled hourly gauges and compared to the corresponding UKCP09 5 km gridded daily

Table 2

Automated rule base definition (QC process step 3) constructed on individual and combined QC flags and based on knowledge of known rainfall processes and errors. If the criteria of any of the rules is fulfilled, the suspect value(s) are treated as erroneous and removed from the record and replaced with a missing data value (–999). QC flags which are applied in step 2 are defined in Table 1.

Rules	QC flag(s) indicating erroneous data	Rule base description	Number of hours removed by rule
R1	QC1, or QC3 flags indicating threshold exceedance of at least 20%	Large hourly or daily values	305
R2	QC1, threshold exceedance of < 20% following at least 12 h of missing values	Large accumulated values	17
R3	QC1 and any other QC flag	Large value and flagged by another QC test	15
R4	QC2	Large “winter” totals following at least 12 h of missing values (e.g. possible frozen gauges)	36
R5	QC2	Large “winter” totals within 24 h of any other flag (possible frozen gauges – successive melting or accumulation causing gauge problems)	89
R6	QC2	Three or more flags in a month suggests persistent problems in a gauge – treat period from first to last instance as suspect	7285
R7	QC2 and QC4 (0900 & 1200)	Accumulations – large “winter” totals that occur at 0900 or 1200 with 23 preceding dry hours	23
R8	QC4 (0900 & 1200) if there is another QC4 flag within 7 days	Suspected accumulation. Remove all values with this flag along with preceding 23 dry hours	171,442
R9	QC5	where > 4 consecutive days are flagged these are eliminated along with intervening dry hours	349,464
R10	QC5	where 3 or 4 consecutive days are flagged these are eliminated only if there are any values exceeding a trace threshold (0.5 mm) along with intervening dry hours	113,664
R11	QC5 and QC4	Where 2 consecutive days are flagged these are only eliminated if QC4 (0900 or 1200) is flagged as being at least 2x mean monthly wet day amount on both days. Intervening dry hours are also treated as suspect.	1008
R12	QC6 and QC3 (any flag)	Monthly accumulations where daily record is exceeded	15,926
R13	QC6	Monthly accumulations where flag occurs in consecutive months	7308
R14	QC11 (45d threshold)	Very long dry spells – and values (potential accumulations) after termination	6,180,498
R15	QC10	Long dry spells occurring at the start or at end of the gauge record	4,349,185
R16	QC8	All frequent tips	380
R17	QC7	All consecutive duplicate values above threshold	295
R18	QC9	All manually identified errors	450,746
R19	QC12, QC13	Neighbourhood dry spell flag	2,573,089
R20	QC14	Neighbourhood high value flag, in winter half year (NDJFMA) only	930,900

observed precipitation time series for the period 1990–2007 (Perry et al., 2009). We used the UKCP09 dataset for validation as the CEH-GEAR dataset was used in the QC process to flag suspicious data (QC9). Although UKCP09 and CEH-GEAR use different interpolation methods, they are still highly correlated. Indeed, when the 24 h accumulation for each CEH-GEAR 1 km time series is compared with its corresponding 5 km grid square, the Spearman's rank and Pearson correlation coefficients range from 0.9 to 1. As the main applications of the gridded hourly data likely relate to the occurrence and intensity of extreme events, this dataset was used to validate the rule base for days which contained very high wet values (wet hour Q99 of the original hourly records) as well as for dry spells longer than 20 days.

Fig. 3 shows the different 24 h Q99 event types highlighted by the validation process. Event type A is where a 24 h Q99 value is found in the gauge data and a wet event is found in the gridded data. As the two values coincide, it is likely that the event did occur. If the rule base excludes this value, we consider it ‘incorrect’. Event type B occurs when there is high rainfall in the gauge data but not in the gridded dataset. For this type of event, if the rule base excludes this high value, we consider it ‘correct’ under the assumption that the gridded dataset is a reliable benchmark. Event type C occurs when there is a relatively high rainfall value observed in the gridded data but not in the gauge data. This type of event would not be excluded by the rule base as the value in the gauge data is low. In order to make this ‘high-low’ assessment of the difference between the two datasets we consider similar events as those where the percentage difference is less than 65% (event type B in Fig. 3, plotting around the 1:1 line on Fig. 4). 65% was chosen as the majority of differences are smaller than this. Conversely, if the difference is large, i.e. > 65%, we would expect that those values should frequently be excluded by the rule base.

In Fig. 4 we compare daily total rainfall (0900 to 0900) encompassing each hourly Q99 event from the hourly gauge records, to the corresponding day and location in the UKCP09 daily dataset. This

shows that as would be hoped, non-excluded points generally cluster around the 1:1 line where the two datasets are in good agreement, whilst a large number of excluded points are characterised by lower correspondence between the two datasets, indicating that generally the rule base is working as intended. In total, 24.3% of type B events were excluded by the rule base in contrast with only 2.0% of type A events.

The largest excluded values are typically eliminated by rules containing threshold checks (as categorised in Tables 1 and 2). A line of events on the ~6:1 line in Fig. 4 are excluded by non-threshold checks which identify unexplained scaling in the magnitude of the hourly record, typically after a period of no data (see Fig. 5) and are likely related to some undocumented gauge malfunction, although this feature did require some manual checking to identify the nature of the problem (QC9). It may be the case that standard statistical tests for break points (e.g. Buishand, 1982; Pettitt, 1979) could instead be used to identify such errors but these also have a number of limitations of their own (Serinaldi et al., 2018). It is also noticeable that a large number of excluded events do lie around the 1:1 line, i.e. potentially ‘incorrectly excluded’ as their daily totals approximately agree. However, further examination indicates that these are typically excluded by rules that apply the daily accumulation flag (QC4), which means that although the daily totals are in broad agreement, the storm shape is wrong in the hourly record – typically as a result of recording the total as a 0900 accumulation, resulting in their exclusion.

Validation by comparison to the daily gridded dataset is a useful method but it is not without limitations, particularly with regard to the lack of commensurability in the spatial representation of the two data sources. No metadata on the locations of the gauges is provided with the UKCP09 5 km gridded daily rainfall dataset, so we may be comparing the hourly data to an interpolated grid square value. We therefore have to be particularly careful when validating in this way as extreme sub-daily rainfall events can be of short duration and limited in their spatial extent and may not be captured in an interpolated grid

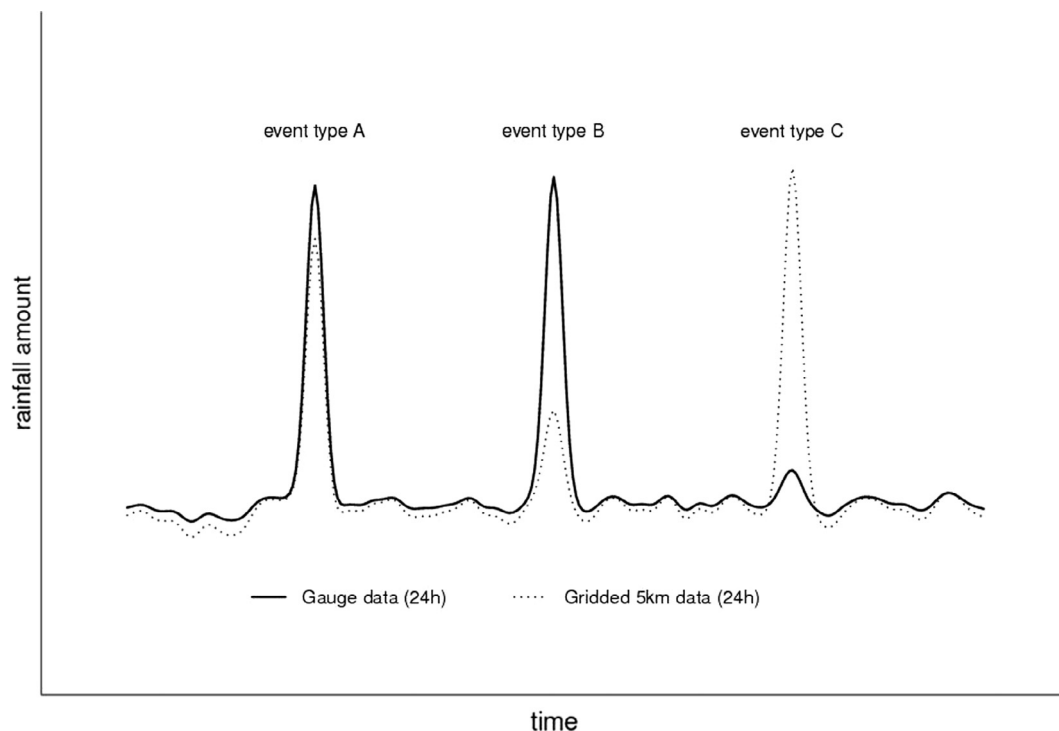


Fig. 3. Schematic of hypothetical scenarios of different large value event types highlighted by the comparison of the quality controlled hourly data (aggregated to daily) to the gridded daily data from the UKCP09 5 km gridded daily dataset as part of the validation process. Event type A is where a Q99 value is found in the gauge data and a wet event is found in the gridded data. Event type B describes high rainfall in the gauge data and low rainfall in the gridded dataset. Event type C describes a relatively high rainfall value observed in the gridded data but not in the gauge data.

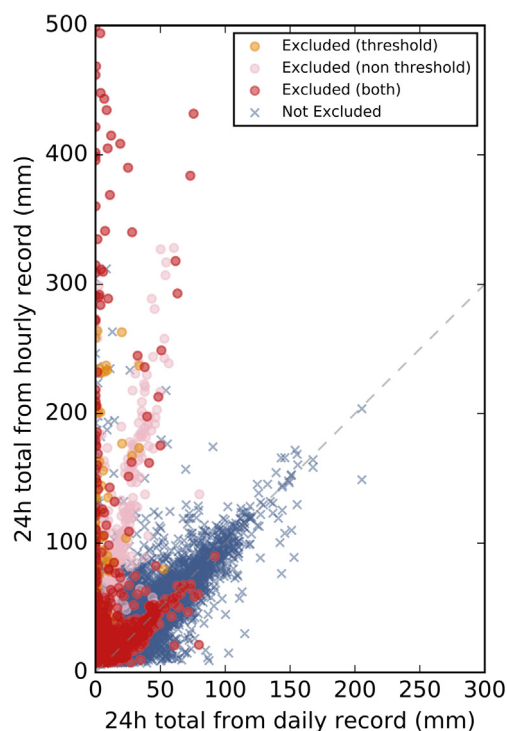


Fig. 4. Comparison of 24 h rainfall totals for hourly Q99 wet hour events (hourly gauge record) with corresponding totals from the daily gridded record. Events are marked as either excluded or not excluded by the rule base. The excluded events are distinguished as either a consequence of rules comprising threshold checks (including neighbourhood checks), non threshold checks or both (See Table 1). The Q99 range of 24 h totals from the hourly record is 6.1–6157 mm.

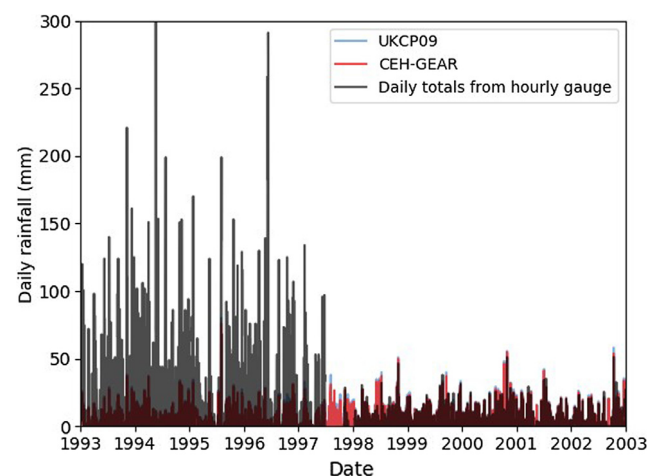


Fig. 5. Example of inter-dataset comparison of accumulated daily rainfall totals for hourly data at Cowbridge, South Wales and corresponding daily totals from the CEH-GEAR and UKCP09 datasets to identify a non-homogeneous time series. These are generally identified by QC9 and removed by R18 (described in Tables 1 and 2).

square, or may be smoothed out by interpolation. This may justify why many type B events were not excluded.

To further check that the rule base was not excluding type A events, we examined the correspondence of our data with known, high-intensity historic events. We took 16 historic extreme rainfall events recorded in ‘Weather’ journal articles and a further 27 from a chronology of severe UK weather events between 1901 and 2008 (Eden, 2008). The location, date, and duration of each event was identified from the literature (see Table S2) and compared to the highest recorded hourly value 24 h either side of the event at the nearest recording gauge. For 12 of these events, no rainfall was recorded in the nearest gauge and for

3, a period of missing data was recorded in the nearest gauge. Of the remaining 28, no events were excluded by the rule base which provides some confidence that it is not excluding real extreme events. This method of validation is somewhat limited as the storm centre and its extent are not reported and so the extreme events may not necessarily show in the nearest hourly gauge record. We also acknowledge that this is a relatively small sample of events compared to the number of gauges but it does provide some confidence that the QC process is not excluding genuine extreme events that were frequently associated with significant impacts.

4.2. Validation of the exclusion of dry spells

Long dry periods are another common error in the hourly gauge records. In particular, long sequences of zero values at the beginning and/or end of records. This may occur when the start or end time of a gauge has been incorrectly reported, its malfunctioning is undocumented, or data values have been incorrectly recorded as zero. Dry spells in the UK are typically defined by a 15-day threshold (Atkinson et al., 1985). For validation of the rule base, we however investigated all sequences of dry days over a relaxed threshold of 20 days or more within the gauge records. For each dry sequence the percentage of wet days and average daily rainfall for the relevant grid cell over the corresponding period in the UKCP09 dataset were calculated. Fig. 6 shows the two types of dry events examined in the validation process. For likely true dry periods (type D events), the % wet days and average daily rainfall in the UKCP09 gridded daily dataset are very low by definition. Type D events are therefore defined as periods that are dry in the gauge data and for which the UKCP09 gridded dataset has a wet-day percentage value of $\leq 20\%$ or an average daily rainfall of ≤ 1 mm. Type E events (likely erroneous), are defined as periods that are dry in the gauge data but for which the UKCP09 gridded dataset has a wet-day percentage value of $> 20\%$ or an average daily rainfall of > 1 mm. In total, 61.4% of type E (erroneous) events were excluded by the rule base, whereas only 3.3% of type D (true) events were excluded when evaluating by wet percentage day. Similarly, 69.3% of type E events were excluded, whilst only 8.0% of type D were when evaluating by

average daily rainfall. The rule base therefore seems to be effective as it is excluding a large percentage of erroneous dry spells whilst only excluding a small percentage of real dry spells. Due to the interpolation of daily rainfall values, however, it is more likely to rain in the gridded record than at the gauge and so, even for true dry sequences, some rain might be expected in the UKCP09 gridded dataset. The percentage wet days and average daily rainfall is much lower for non-excluded events than those that are excluded (Fig. 7) demonstrating that the rule base is excluding mainly erroneous dry spells.

5. Resulting gauge dataset

In total, 3.4% of the hourly data was excluded by the QC process. Fig. 8 shows the improvement over Fig. 2 in r_s and percentage correct statistics after the rule base is applied (mean absolute differences are also shown in Fig. S3). Only 2.5% of gauges still have poor correlation and percentage correct statistics (< 0.8) when compared to the CEH-GEAR daily time series. On further investigation, these gauges are characterised by rainfall values of a reasonable order of magnitude but with long periods of missing data suggesting that the gauge may be faulty over a prolonged period. The absence of particularly high values in these gauges means that the QC tests have not flagged data as suspicious, and therefore such potential discrepancies are best identified through comparison to a high quality reference dataset (e.g. the CEH-GEAR gridded dataset, check gauges etc.). Such gauges would be excluded from some climatological analyses as a consequence of the large percentage of missing data though this is not an important factor for the production of the gridded dataset as the subsequent interpolation procedure (Section 6) accounts for missing periods by using the next nearest gauge. However, as a precaution against using such potentially erroneous data, only gauges with $r_s > 0.8$ and $P00 + P11 > 0.8$ after the QC process were used.

There are a range of rule bases that could be used to determine the exclusion of potentially suspect data flagged by the QC tests outlined in this paper. We therefore examined three other rule bases made up of fewer rules and representing differing levels of ‘severity’. Each of these was validated using the process described above (see Supplementary

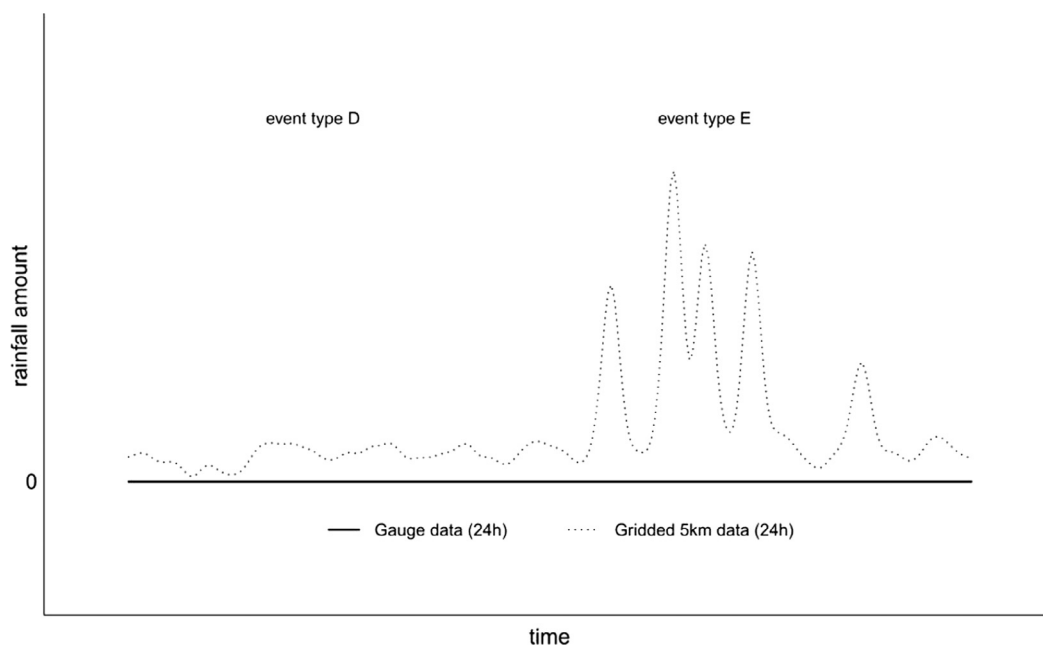


Fig. 6. Schematic of different dry spell event types highlighted by the comparison of the quality controlled hourly data (aggregated to daily) to the gridded daily data from the UKCP09 5 km gridded daily dataset as part of the validation process. Type D events are defined as periods that are dry in the gauge data and for which the UKCP09 gridded dataset has a wet-day percentage value of $\leq 20\%$ or an average daily rainfall of ≤ 1 mm. Type E events (likely erroneous), are defined as periods that are dry in the gauge data but for which the UKCP09 gridded dataset has a wet-day percentage value of $> 20\%$ or an average daily rainfall of > 1 mm.

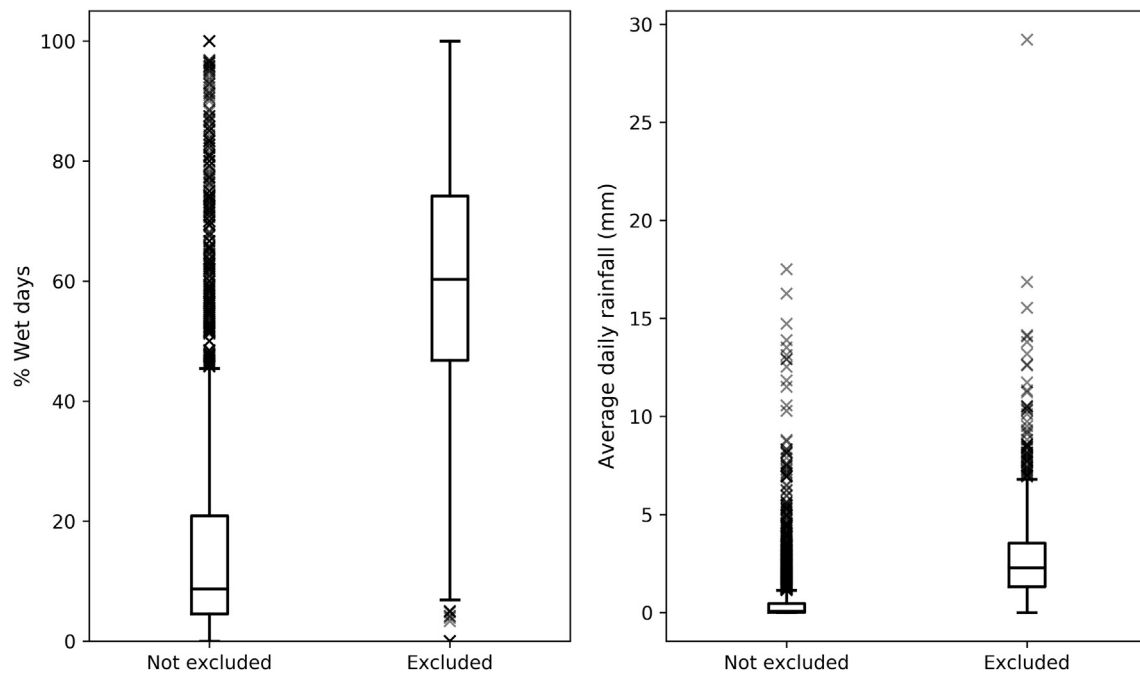


Fig. 7. For each individual dry spell event in the hourly record longer than 20 days, the corresponding time series is found in the UKCP09 gridded dataset. The percentage of days for which it is raining in the gridded dataset (left) and the mean daily rainfall in the gridded dataset (right) for each period is plotted for events excluded by the rule base and not excluded by the rule base.

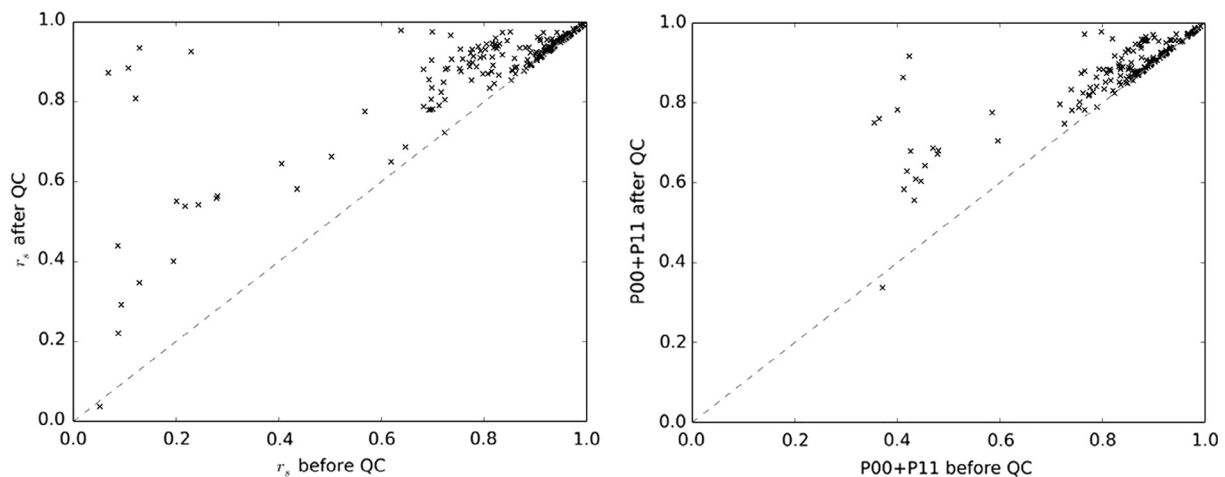


Fig. 8. Comparison of r_s values between each hourly rain gauge (accumulated to 24 h) and the corresponding CEH-GEAR daily time series before and after implementation of the QC process (left), and for P00 + P11 (percentage correct) statistics before and after implementation of QC process (right).

Information, Tables S3 and S4). The selected rule base described in this paper was found to eliminate a relatively large amount of suspect data whilst simultaneously eliminating only a small amount of non-suspect data. It is difficult to quantitatively demonstrate which rule base is 'better' as we do not have a reliable true reference dataset. The rule base presented here representatively codifies the judgments made in manual inspection and was therefore considered to be the most appropriate.

6. Temporal disaggregation

The quality controlled hourly gauge data was used to disaggregate the CEH-GEAR gridded daily rainfall dataset. There are many different interpolation methods available for gridding rainfall data, such as Thiessen, inverse distance weighting, cubic spline, kriging etc. Many studies have compared the relative benefits of each (Contractor et al., 2015; Dunn et al., 2014; Hofstra et al., 2008; Dirks et al. 1998) and

conclude that in areas of high gauge network density, the method selected has little impact. Given the relatively high rainfall gauge network density in the UK, a simple nearest neighbour interpolation without height correction was used to preserve a real storm shape for every grid square. This was considered to be beneficial as it will preserve extreme hourly rainfall intensities whereas other interpolation methods will smooth these extremes out. A limitation of this approach is that convective events can be very small and therefore nearest neighbour may sometimes represent a convective storm over too large an area. However, as it is the hourly rainfall fractions that are interpolated here, the actual rainfall total is modulated by the daily rainfall dataset, which is smoothed, meaning that this effect is reduced. This methodological choice is supported by similar applications in Li et al. (2018), and Choi et al. (2008).

For each day a subset of hourly gauges was selected, using only those where the record for that day was complete. The number of gauges used on each day was therefore variable as some were excluded

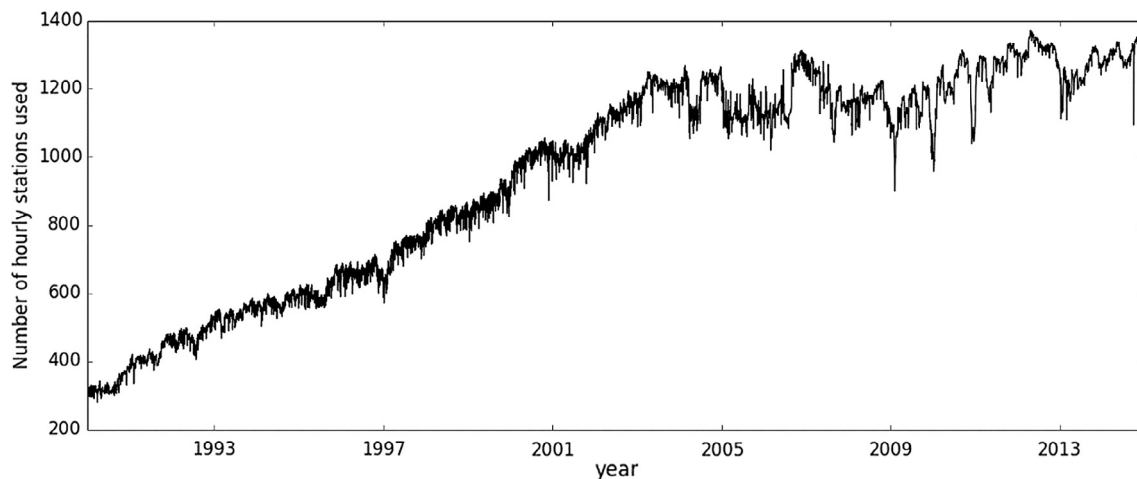


Fig. 9. Number of stations used for disaggregation each day. The number used increases from 295 at the beginning of the dataset and peaks at 1372 gauges in April 2012.

because they contained missing data. Therefore, the number of gauges used on a given day ranged from 295 to 1372 (see Fig. 9). For each hour, the gauge data was interpolated onto a 1 km grid using nearest neighbour interpolation that, for each grid square, assigns it the nearest station hourly rainfall value as a fraction of the station's daily total (Isaaks and Srivastava, 1989). If the nearest station was over 50 km away, or if there was rain in the daily dataset but not in the nearest hourly gauge, the station value was not used and an average storm shape was used instead. The gridded hourly fractions were then multiplied by the daily rainfall from the CEH-GEAR dataset. This preserved the daily totals from CEH-GEAR whilst maintaining the nearest recorded storm shape.

Fig. 10 shows the average storm shape given a daily rainfall total range, which changes according to season, and was constructed from all the available gauges (see Fig. S2 for the full distribution of storm profiles). In winter (November–April) the storms are typically longer in

duration and less intense whilst in summer (May–October) they are shorter and more intense. This ‘average storm’ is set to begin in the gridded dataset at 0900 whenever it is applied. This represents a more realistic approach to disaggregation than has been used elsewhere, when accumulated rainfall totals have been distributed equally across the 24 h period (e.g. Parkes et al., 2013) and could potentially be improved upon by setting the peak of the design storm to coincide with a seasonal likely wettest hour. The design storms could also be calculated regionally, or alternatively a weather generator could be used to infill missing data.

7. Reliability of the gridded hourly dataset

The reliability of CEH-GEAR1hr is dependent upon that of the daily totals from CEH-GEAR (see Keller et al. 2015), and of the hourly disaggregations. The disaggregation error has several components:

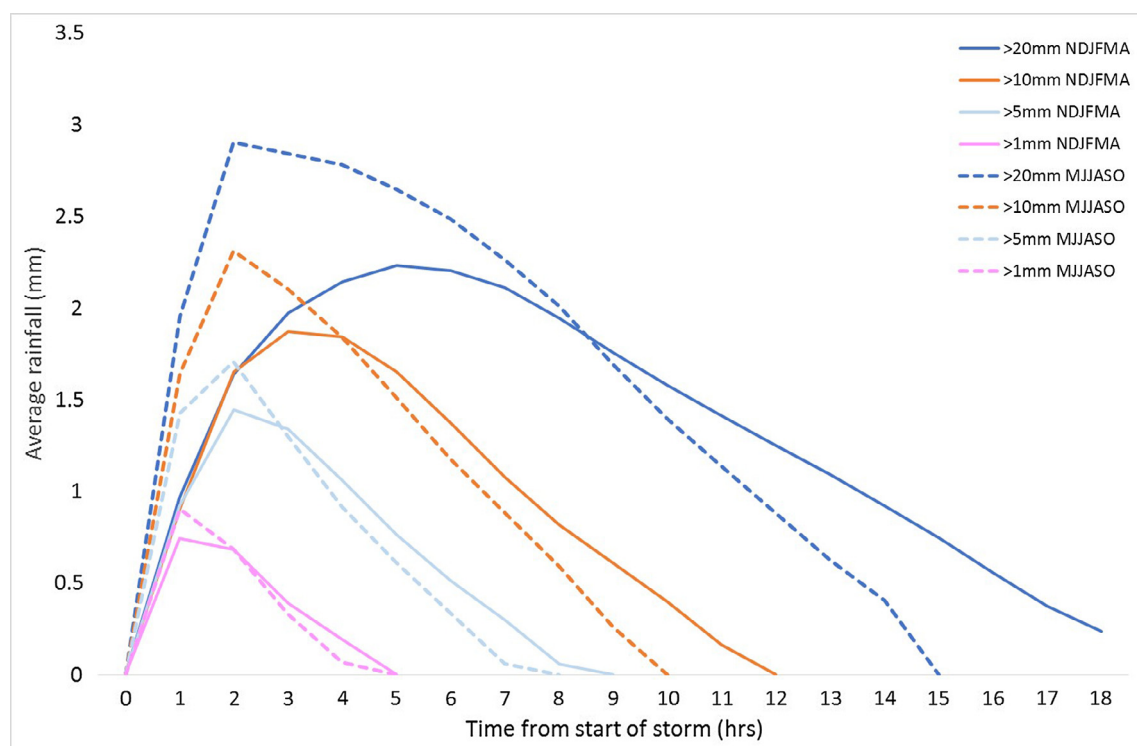


Fig. 10. Average storm shapes for different total daily rainfall thresholds. Separate profiles are provided for the winter (Nov–April) and summer (May–Oct) half years.

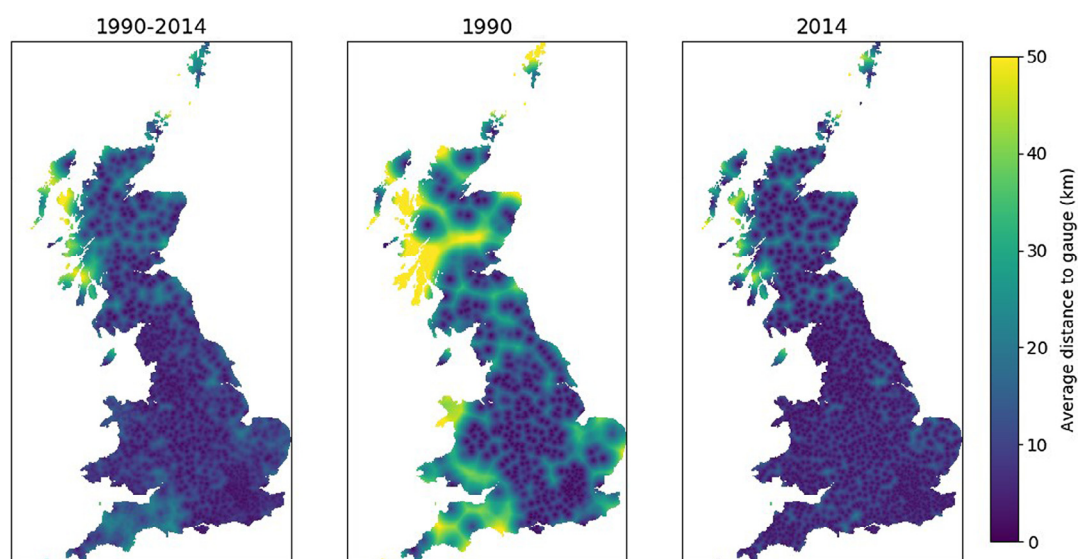


Fig. 11. Average distance to an hourly gauge for each grid square over the period 1990–2014 and for the years 1990 and 2014 respectively.

measurement error of the gauge record, error associated with the distance to the nearest gauge and the error associated with using statistical disaggregation. The measurement error has been reduced as much as possible through QC; however, some errors will remain in the gauge data, particularly those associated with wind-induced under-catch and evaporation errors.

The metadata associated with the dataset includes the distance to the gauge used for each grid square for every day. The mean distance to a gauge over the whole period is 11.3 km and the maximum distance is 97.7 km for the west coast of Scotland. Most of the country is reasonably well covered by rain gauges, with the exception of Scotland and the south west of England, which is reflected in these areas having a higher average distance to gauge (Fig. 11). In the case of the latter, this is likely a result of the later instrumentation with gauges by the Environment Agency in south west England compared with other parts of the country. Fig. 12 shows the increasing error with distance between the temporal patterns (the fraction of daily rainfall falling in each hour) of a gauge and its nearest neighbour. This information, together with the distance to the gauge used for disaggregation provided with the hourly rainfall grids, gives users the information to decide if parts of the CEH-GEAR1hr estimates are suitable for their needs. The low gauge density in some areas also means that errors are likely to occur from under-sampling of orographically enhanced rainfall and localised

convective storms.

The metadata also describes whether or not statistical disaggregation was used for each day. As noted above, where a grid square is greater than 50 km from a gauge, or when there is rainfall in the daily record but not in the hourly record, statistical disaggregation is used. Statistical disaggregation due to zero rainfall in the hourly record is used for 0–26% of grid squares on any given day over the whole 1990–2014 period, whereas statistical disaggregation due to the grid square being over 50 km away from a gauge was used for 0–3% of grid squares over this time period. Instances of the latter use of statistical disaggregation are rare after 2000 when the final regions of the EA network were gauged and so the average distance between gauges in data sparse areas decreased. Whilst the percentage of grid squares using statistical disaggregation due to zero rainfall in the hourly record fluctuates greatly over the record, 97% of the time this is used to disaggregate a daily total of less than 1 mm of rainfall. The statistical disaggregation is therefore unlikely to have a large impact on extreme values in the dataset or on subsequent hydrological model simulations.

8. Evaluation of the hourly gridded dataset

CEH-GEAR1hr has statistical properties of interest that are generally consistent with previous gauge-derived national scale estimates

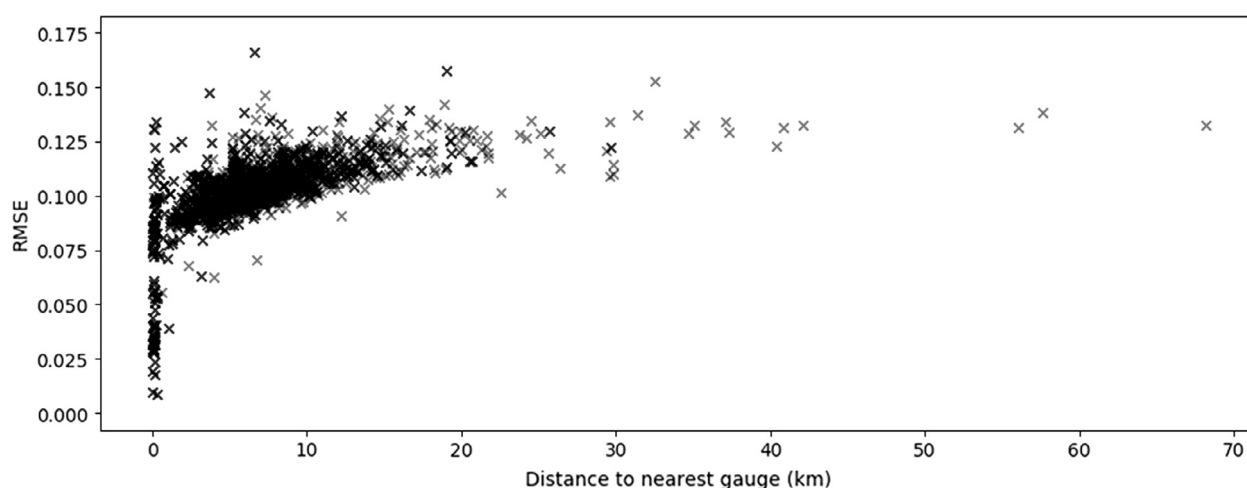


Fig. 12. Root mean square error (RMSE) of the hourly fraction of daily rainfall represented as a function of the distance to the closest rain gauge.

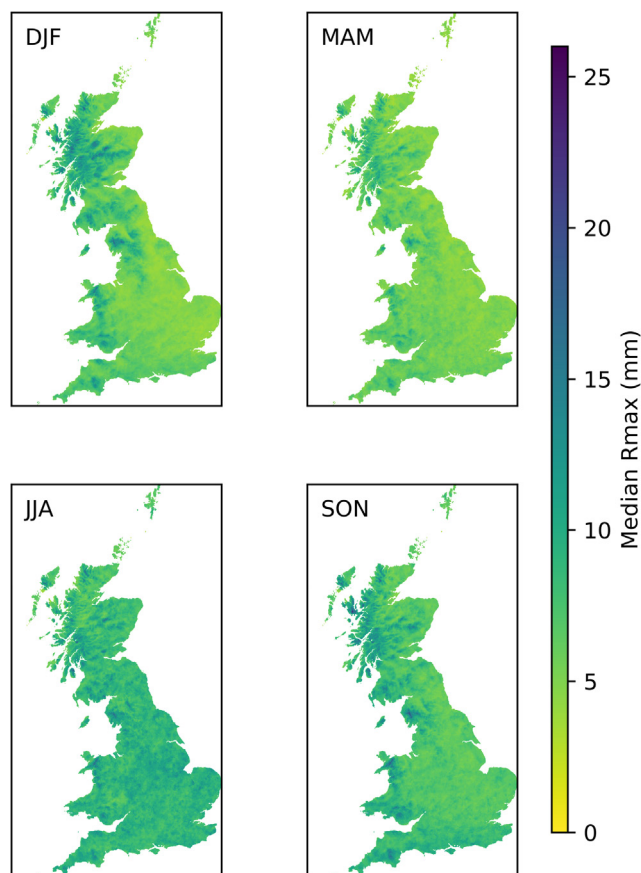


Fig. 13. The median seasonal maximum 1 h rainfall (Rmax) for the period 1990–2014 calculated from CEH-GEAR1hr.

(Blenkinsop et al., 2017). Throughout most of the year the median of the seasonal maximum 1 hr rainfall (Rmax) is highest in the mountainous regions and the west of Britain (Fig. 13) whilst in the east and lowland areas, Rmax is highest in summer (JJA) and autumn (SON). The spatial pattern of Rmax is also much less coherent and less clearly defined by orography in summer which might be expected from the increased dominance of convective rainfall at this time of year and is consistent with the corresponding results presented in Blenkinsop et al. (2017). Summer also sees the lowest number of wet hours across the country and winter the greatest (Fig. 14), with the influence of topography on this variable evident throughout the year. These patterns are, as expected, broadly consistent with those of daily rainfall occurrence rates (Jenkins et al., 2008). By using both gridded daily data and hourly gauges, the new dataset provides additional spatial detail not present in previous maps. A more detailed evaluation of the dataset is beyond the scope of this paper but will be provided in a subsequent publication.

9. Discussion and conclusions

A 1 km gridded hourly rainfall dataset has been created for Great Britain using data from over 1900 quality controlled gauges for the period 1990–2014. We extended and automated a quality control procedure (essential for such a large number of gauges) to expand the availability of the hourly rain gauge dataset accumulated by Blenkinsop et al. (2017). We allocated a series of QC flags that identified *potentially suspect* values and then constructed a set of rules that apply either single or combined QC flags based on knowledge gained from the initial dataset to exclude *likely erroneous* rainfall amounts, accumulations and dry spells. A two-tiered validation approach at the daily and hourly timescales indicates that spurious extreme values have been excluded from the resultant dataset, while assumed legitimate values have been

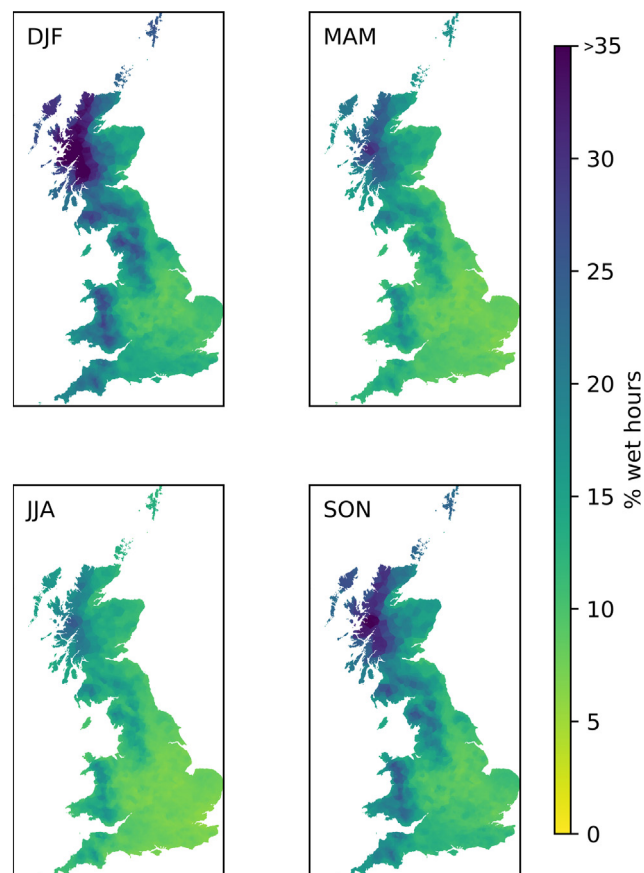


Fig. 14. The percentage of wet hours per season for the period 1990–2014 calculated from CEH-GEAR1hr.

preserved. Most quality-controlled gauges are found to be highly correlated with the UKCP09 observed 5 km dataset at the daily timescale and show a high degree of concordance in terms of rainfall occurrence. The resulting gauge density used in this product varies over time and space, being particularly sparse in Scotland and the southwest of England prior to 2000, and this should be borne in mind when using the dataset.

A nearest neighbour interpolation scheme was used to provide gridded hourly rainfall values at a resolution of 1 km. These data were then used to temporally disaggregate the existing CEH-GEAR daily gridded dataset to produce an hourly dataset with consistent daily totals. Consistency with the existing CEH-GEAR dataset is an important feature of the new hourly dataset as it will be made freely available, hosted and updated alongside the CEH-GEAR product at <http://eidc.ceh.ac.uk>.

This new dataset will be a valuable resource for hydrologists, climate scientists and the broader community wishing to assess current exposure to intense rainfall. There are few national datasets available at a sub-daily time-step yet many hydrological applications require, or can be improved by, a higher temporal resolution of rainfall data especially for smaller, rapidly responding, catchments (e.g. Archer et al., 2016). However, we would recommend that the gridded hourly data should not be used for trend analyses due to the short length of the dataset, potential gauge level inhomogeneities and the temporal variation in gauges used in the disaggregation. Kendon et al. (2018) examined UK gauges of at least 13 years duration and subsequently corrected for inhomogeneities arising from changes in measurement resolution. This was found to have a limited effect on extreme values but can have a significant impact on mean intensities and rainfall occurrence statistics. The lack of additional metadata available makes the identification and attribution of other potential inhomogeneities challenging.

If only gauge data is required, the QC procedure presented here can be used to improve the consistency of data by trying to eliminate erroneous extremes whilst maintaining genuine values. The QC framework here will therefore also be implemented in the global effort to gather sub-daily rainfall data currently being undertaken by the INTENSE (INTElligent use of climate models for adaptation to non-Stationary hydrological Extremes) project which forms part of the World Climate Research Programme's Global Energy and Water EXchanges (GEWEX) Grand Challenge on Extremes. This has identified the importance of developing new and novel QC methods at different timescales and for different locations (Alexander et al., 2016). A key challenge will be to adapt the tests and rules developed here so that they are applicable to different climate regimes and operating practices.

A by-product from the generation of the gridded dataset is an additional dataset comprising the individual QC flags (Table 1) associated with each hourly value for each gauge. This metadata can be obtained alongside the rule base code to allow users who are licensed to use the original gauge data to apply a bespoke rule base to the gauge data. This creates a community resource where improvements to the database can be made and shared for different types of analyses. One future piece of work could be to explore the uncertainty associated with the implementation of the rule base or, more generally, the quantification of uncertainty within the gridded hourly precipitation product. The importance of adequately assessing uncertainty within model predictions and observations that drive them has been highlighted by a vast array of authors (e.g. Beven, 2002; Pappenberger et al., 2006, 2008; Di Baldassarre et al., 2010; McMillan et al., 2012). Within the field of hydrology, the rainfall data used to force a model prediction is a key source of error, and therefore, the generation of a probabilistic rainfall product would be a valuable next step (Ahrens and Jaun 2007; Sideris et al. 2014). Corresponding products are currently being developed by the reanalysis community (Bach et al., 2016). Some key prerequisites for generating a probabilistic rainfall product would be to gain a better understanding of the uncertainties associated with the hourly gauge data (for example undercatch), the QC process and the interpolation method, but ultimately this would also need to be explored at the daily level in the original CEH-GEAR product.

Further potential exists for novel methods to create merged products that take advantage of the spatial detail offered by other sources of data such as radar. Jewell and Gaussiat (2015) examined several schemes to merge radar with TBR data over England and Wales, all of which produced a merged product that was superior to the individual data sources. Such products have significant potential to support flood forecasting and may provide improved calibration of hydrological models (Parkes et al., 2013). As a next step, we intend to provide added value to the dataset by conducting an assessment of its key characteristics with respect to flooding from intense rainfall and developing freely available gridded datasets of useful indices such as extreme percentiles and peaks over threshold as well as the provision of Intensity-Duration-Frequency curves. We will also evaluate the benefit of using the hourly gridded dataset in hydrological modelling applications.

10. Declarations of interest

None.

Acknowledgements

This research forms part of the SINATRA project which is supported by the United Kingdom NERC Flooding from Intense Rainfall programme (grant NE/K00896X/1) and an associated knowledge exchange award. It was also funded as part of the CONVEX project which was supported by the United Kingdom NERC Changing Water Cycle programme (grant NE/I006680/1) and the INTENSE project through the European Research Council (grant ERC-2013-CoG-617329). Hayley

Fowler is funded by the Wolfson Foundation and the Royal Society as a Royal Society Wolfson Research Merit Award (WM140025) holder. Paul Bates was supported by a Research Fellowship from the Leverhulme Trust and a Royal Society Wolfson Research Merit Award. We would also like to thank Carol Langley of the UK Environment Agency, Michelle Lewis of NRW and Grant Kennedy of SEPA for the provision of precipitation data used in this paper.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.jhydrol.2018.07.034>.

References

- Abbott, P.F., 1986. Guidelines on the Quality Control of Surface Climatological Data. World Meteorological Organization, Geneva, Switzerland (WMO/TD-No. 111).
- Ahrens, B., Jaun, S., 2007. On evaluation of ensemble precipitation forecasts with observation-based ensembles. *Adv. Geosci.* 10, 139–144.
- Alexander, L., Zhang, X., Hegerl, G., Seneviratne, S., Behrangi, A., Fischer, E., Martius, O., Otto, F., Sillmann, J. and Vautard, R., 2016. Implementation plan for WCRP grand challenge on understanding and predicting weather and climate extremes. World Climate Research Programme Report: Geneva, Switzerland. https://www.wcrp-climate.org/images/documents/grand_challenges/WCRP_Grand_Challenge_Extremes_Implementation_Plan_v20160708.pdf (accessed 5 June 2018).
- Archer, D.R., Parkin, G., Fowler, H.J., 2016. Assessing long term flash flooding frequency using historical information. *Hydrol. Res.* 48 (1), 1–16. <https://doi.org/10.2166/nh.2016.031>.
- Atkinson, B.W., Penning-Rowsell, E.C., Porter, D.J., 1985. Reviews of United Kingdom Statistical Sources. In: *Weather and Water*. Royal Statistical Society and Economic and Social Council. Pergamon Press, pp. 266.
- Bach, L., Schraff, C., Keller, J.D., Hense, A., 2016. Towards a probabilistic regional reanalysis system for Europe: evaluation of precipitation from experiments. *Tellus A: Dyn. Meteorol. Oceanogr.* 68, 32209. <https://doi.org/10.3402/tellusa.v68.32209>.
- Bastola, S., Misra, V., 2013. Sensitivity of hydrological simulations of southeastern United States watersheds to temporal aggregation of rainfall. *J. Hydrometeorol.* 14 (4), 1334–1344. <https://doi.org/10.1175/JHM-D-12-096.1>.
- Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., Ziese, M., 2013. A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901-present. *Earth Syst. Sci.* 5, 71–99. <https://doi.org/10.5194/essd-5-71-2013>.
- Bell, V.A., Kay, A.L., Jones, R.G., Moore, R.J., 2007. Development of a high resolution grid-based river flow model for use with regional climate model output. *Hydrol. Earth Syst. Sci.* 11, 532–549. <https://doi.org/10.5194/hess-11-532-2007>.
- Beven, K., 2002. Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. *Hydrol. Process.* 16, 189–206. <https://doi.org/10.1002/hyp.343>.
- Blenkinsop, S., Fowler, H.J., Dubus, I.G., Nolan, B.T., Hollis, J.M., 2008. Developing climatic scenarios for pesticide fate modelling in Europe. *Environ. Poll.* 154, 219–231.
- Blenkinsop, S., Lewis, E., Chan, S.C., Fowler, H.J., 2017. An hourly precipitation dataset and climatology of extremes for the UK. *Int. J. Climatol.* 37, 722–740. <https://doi.org/10.1002/joc.4735>.
- Borsche, M., Kaiser-Weiss, A.K., Undén, P., Kaspar, F., 2015. Methodologies to characterize uncertainties in regional reanalyses. *Adv. Sci. Res.* 12, 207–218. <https://doi.org/10.5194/asr-12-207-2015>.
- Buishand T.A., 1982. Some methods for testing the homogeneity of rainfall records. KNMI Scientific Report WR 81-7, De Bilt, The Netherlands.
- Chan, S.C., Kendon, E.J., Fowler, H.J., Blenkinsop, S., Roberts, N., Ferro, C.A.T., 2014. The value of high-resolution Met Office regional climate models in the simulation of multi-hourly precipitation extremes. *J. Clim.* 27, 6155–6174.
- Choi, J., Socolofsky, S.A., Olivera, F., 2008. Hourly disaggregation of daily rainfall in Texas using measured hourly precipitation at other locations. *J. Hydrol. Eng.* 13 (6), 476–487.
- Cole, S., Moore, R., 2008. Hydrological modelling using rain-gauge and radar-based estimators of areal rainfall. *J. Hydrol.* 358, 159–181.
- Cole, S., Moore, R., 2009. Distributed hydrological modelling using weather radar in gauged and ungauged basins. *Adv. Water Resour.* 32, 1107–1120.
- Collier, C.G., 1989. *Applications of Weather Radar Systems*. Ellis Horwood, Chichester, UK.
- Compo, G.P., Whitaker, J.S., Sardeshmukh, P.D., Matsui, N., Allan, R.J., Yin, X., Gleason, B.E., Vose, R.S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R.I., Grant, A.N., Groisman, P.Y., Jones, P.D., Kruk, M., Kruger, A.C., Marshall, G.J., Maugeri, M., Mok, H.Y., Nordli, Ø., Ross, T.F., Trigo, R.M., Wang, X.L., Woodruff, S.D., Worley, S.J., 2011. The twentieth century reanalysis project. *Quarterly J. Roy. Meteorol. Soc.* 137, 1–28. <https://doi.org/10.1002/qj.776>.
- Contractor, S., Alexander, L.V., Donat, M.G., Herold, N., 2015. How well do gridded datasets of observed daily precipitation compare over Australia? *Adv. Meteorol.* 2015. <https://doi.org/10.1155/2015/325718>.
- Cosgrove, B.A., Lohmann, D., Mitchell, K.E., Houser, P.R., Wood, E.F., Schaake, J.C.,

- Robock, A., Marshall, C., Sheffield, J., Duan, Q., Luo, L., Higgins, R.W., Pinker, R.T., Tarpley, J.D., Meng, J., 2003. Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *J. Geophys. Res.* 108, 8842. <https://doi.org/10.1029/2002JD003118>, D22.
- Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hölml, E.V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q.J.R. Meteorol. Soc.* 137, 553–597. <https://doi.org/10.1002/qj.828>.
- Di Baldassarre, G., Schumann, G., Bates, P.D., Freer, J.E., Beven, K.J., 2010. Flood-plain mapping: a critical discussion of deterministic and probabilistic approaches. *Hydrol. Sci. J.* 55 (3), 364–376. <https://doi.org/10.1080/02626661003683389>.
- Dirks, K.N., Hay, J.E., Stow, C.D., Harris, D., 1998. High-resolution studies of rainfall on Norfolk Island: Part II: Interpolation of rainfall data. *J. Hydrol.* 208 (3), 187–193. [https://doi.org/10.1016/S0022-1694\(98\)00155-3](https://doi.org/10.1016/S0022-1694(98)00155-3).
- Doe, R.K., 2004. Extreme precipitation and run-off induced flash flooding at Boscawle, Cornwall, UK – 16 August 2004. *J. Meteorol.* 29, 319–333.
- Dunn, R.J.H., Donat, M.G., Alexander, L.V., 2014. Investigating uncertainties in global gridded datasets of climate extremes. *Clim. Past* 10 (6), 2171.
- Eden, P., 2008. Great British Weather Disasters. Continuum, London, UK.
- Eischeid, J.K., Bruce Baker, C., Karl, T.R., Diaz, H.F., 1995. The quality control of long-term climatological data using objective data analysis. *J. Appl. Meteor.* 34, 2787–2795. [https://doi.org/10.1175/1520-0450\(1995\)034<2787:TQCOLT>2.0.CO;2](https://doi.org/10.1175/1520-0450(1995)034<2787:TQCOLT>2.0.CO;2).
- Finnerty, B.D., Smith, M.B., Seo, D.J., Koren, V., Moglen, G.E., 1997. Space-time scale sensitivity of the Sacramento model to radar-gage precipitation inputs. *J. Hydrol.* 203 (1–4), 21–38.
- Higgins, R.W., Janowiak, J.E., Yao, Y.-P., 1996. A gridded hourly precipitation data base for the United States (1963–1993). In: NCEP/Climate Prediction Center Atlas 1. National Centers for Environmental Prediction, pp. 46.
- Hofstra, N., Haylock, M., New, M., Jones, P., Frei, C., 2008. Comparison of six methods for the interpolation of daily, European climate data. *J. Geophys. Res.* 113, D21110. <https://doi.org/10.1029/2008JD010100>.
- Isaaks, E.H., Srivastava, R.M., 1989. Applied Geostatistics. Oxford University Press, New York.
- Isotta, F.A., Vogel, R., Frei, C., 2015. Evaluation of European regional reanalyses and downscalings for precipitation in the Alpine region. *Meteorologische Zeitschrift* 24, 15–37.
- Jenkins, G.J., Perry, M.C., Prior, M.J., 2008. The Climate of the United Kingdom and Recent Trends. Met Office Hadley Centre, Exeter, UK.
- Jewell, S.A., Gaussiat, N., 2015. An assessment of kriging-based rain-gauge–radar merging techniques. *Q.J.R. Meteorol. Soc.* 141, 2300–2313. <https://doi.org/10.1002/qj.2522>.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., Joseph, D., 1996. The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Am. Meteorol. Soc.* 77, 437–472. [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- Keller, V.D.J., Tanguy, M., Prosdoci, I., Terry, J.A., Hitt, O., Cole, S.J., Fry, M., Morris, D.G., Dixon, H., 2015. CEH-GEAR: 1 km resolution daily and monthly areal rainfall estimates for the UK for hydrological and other applications. *Earth Syst. Sci. Data* 7, 143–155.
- Kendon, E.J., Roberts, N.M., Senior, C.A., Roberts, M.J., 2012. Realism of rainfall in a very high resolution regional climate model. *J. Clim.* 25, 5791–5806.
- Kendon, E.J., Roberts, N.M., Fowler, H.J., Roberts, M.J., Chan, S.C., Senior, C.A., 2014. Heavier summer downpours with climate change revealed by weather forecast resolution model. *Nat. Clim. Change* 4, 570–576.
- Kendon, E.J., Blenkinsop, S., Fowler, H.J., 2018. When will we detect changes in short-duration precipitation extremes? *J. Clim.* 31, 2945–2964. <https://doi.org/10.1175/JCLI-D-17-0435.1>.
- Li, X., Meshgi, A., Wang, X., Zhang, J., Tay, S.H., Pijcke, G., Manocha, N., Ong, M., Nguyen, M.T., Babovic, V., 2018. Three resampling approaches based on method of fragments for daily-to-subdaily precipitation disaggregation. *Int. J. Climatol.* 38, 1119–1138.
- Luo, Y., Qian, W., Zhang, R., Zhang, D., 2013. Gridded hourly precipitation analysis from high-density rain gauge network over the Yangtze-Huai rivers basin during the 2007 Mei-Yu season and comparison with CMORPH. *J. Hydrometeorol.* 14, 1243–1258. <https://doi.org/10.1175/JHM-D-12-0133.1>.
- McMillan, H., Krueger, T., Freer, J., 2012. Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. *Hydrol. Process.* 26, 4078–4111. <https://doi.org/10.1002/hyp.9384>.
- Met Office, 2012. Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853–current). NCAS British Atmospheric Data Centre, last accessed 14 June 2015. <http://catalogue.ceda.ac.uk/uuid/220a65615218d5c9cc9e4785a3234bdo>.
- Pappenberger, F., Bartholmes, J., Thielen, J., Cloke, H.L., Buizza, R., de Roo, A., 2008. New dimensions in early flood warning across the globe using grand-ensemble weather predictions. *Geophys. Res. Lett.* 35, L10404. <https://doi.org/10.1029/2008GL033837>.
- Pappenberger, F., Beven, K.J., 2006. Ignorance is bliss: or seven reasons not to use uncertainty analysis. *Water Resour. Res.* 42, W05302. <https://doi.org/10.1029/2005WR004820>.
- Parkes, B.L., Wetterhall, F., Pappenberger, F., He, Y., Malamud, B.D., Cloke, H.L., 2013. Assessment of a 1-hour gridded precipitation dataset to drive a hydrological model: a case study of the summer 2007 floods in the Upper Severn, UK. *Hydrol. Res.* 44 (1), 89–105. <https://doi.org/10.2166/nh.2011.025>.
- Paulat, M., Frei, C., Hagen, M., Wernli, H., 2008. A gridded dataset of hourly precipitation in Germany: its construction, climatology and application. *Meteorologische Zeitschrift* 17 (6), 719–732. <https://doi.org/10.1127/0941-2948/2008/0332>.
- Perry, M.C., Hollis, D., Elms, M., 2009. The generation of daily gridded datasets of temperature and rainfall for the UK. Met Office National Climate Information Centre Climate Memorandum No. 24, Met Office, UK.
- Pettitt, A.N., 1979. A non-parametric approach to the change-point detection. *Appl. Statist.* 28, 126–135.
- Prein, A.F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., Keller, M., Tölle, M., Gutjahr, O., Feser, F., Brisson, E., Kollet, S., Schmidli, J., van Lipzig, N.P.M., Leung, R., 2015. A review on regional convection-permitting climate modeling: demonstrations, prospects, and challenges. *Rev. Geophys.* 53, 323–361.
- Rhodes, R.L., Shaffrey, L.C., Gray, S.L., 2015. Can reanalyses represent extreme precipitation over England and Wales? *Q.J.R. Meteorol. Soc.* 141, 1114–1120. <https://doi.org/10.1002/qj.2418>.
- Sciuto, G., Bonaccorso, B., Cancelliere, A., Rossi, G., 2009. Quality control of daily rainfall data with neural networks. *J. Hydrol.* 364, 13–22.
- Serinaldi, F., Kilsby, C.G., Lombardo, F., 2018. Untenable nonstationarity: an assessment of the fitness for purpose of trend tests in hydrology. *Adv. Water Resour.* 111, 132–155.
- Shen, Y., Zhao, P., Pan, Y., Yu, J., 2014. A high spatiotemporal gauge-satellite merged precipitation analysis over China. *J. Geophys. Res.* 119, 3063–3075.
- Sideris, I.V., Gabella, M., Erdin, R., Germann, U., 2014. Real-time radar-raingauge merging using spatiotemporal co-kriging with external drift in the alpine terrain of Switzerland. *Quart. J. Royal Meteorol. Soc.* 140, 1197–1211.
- Simpson, I.R., Jones, P.D., 2014. Analysis of UK precipitation extremes derived from Met Office gridded data. *Int. J. Climatol.* 34, 2438–2449. <https://doi.org/10.1002/joc.3850>.
- Tanguy, M., Dixon, H., Prosdoci, I., Morris, D.G., Keller, V.D.J., 2016. Gridded estimates of daily and monthly areal rainfall for the United Kingdom (1890–2015) [CEH-GEAR]. NERC Environ. Inform. Data Centre. <https://doi.org/10.5285/33604ea0-c238-4488-813d-0ad9ab7c51ca>.
- Upton, G.J.G., Rahimi, A.R., 2003. On-line detection of errors in tipping-bucket rain-gauges. *J. Hydrol.* 278, 197–212.
- Villarini, G., Krajewski, W.F., 2010. Review of the different sources of uncertainty in single polarization radar-based estimates of rainfall. *Surv. Geophys.* 31, 107. <https://doi.org/10.1007/s10712-009-9079-x>.
- Vormoor, K., Skaugen, T., 2013. Temporal disaggregation of daily temperature and precipitation grid data for Norway. *J. Hydrometeorol.* 14, 989–999. <https://doi.org/10.1175/JHM-D-12-0139.1>.
- Westra, S., Fowler, H.J., Evans, J.P., Alexander, L.V., Berg, P., Johnson, F., Kendon, E.J., Lenderink, G., Roberts, N.M., 2014. Future changes to the intensity and frequency of short-duration extreme rainfall. *Rev. Geophys.* 52, 522–555.
- Wilks, D.S., 2006. Statistical Methods in the Atmospheric Sciences, second ed. Academic Press, San Diego, CA.
- Wüest, M., Frei, C., Althoff, A., Hagen, M., Litschi, M., Schär, C., 2010. A gridded hourly precipitation dataset for Switzerland using rain-gauge analysis and radar-based disaggregation. *Int. J. Climatol.* 30, 1764–1775. <https://doi.org/10.1002/joc.2025>.
- Yang, Y., Wang, G., Wang, L., Yu, J., Xu, Z., 2014. Evaluation of gridded precipitation data for driving SWAT model in area upstream of three gorges reservoir. *PLoS One* 9 (11), e112725. <https://doi.org/10.1371/journal.pone.0112725>.
- Yoo, C., Ha, E., 2007. Effect of zero measurements on the spatial correlation structure of rainfall. *Stochast. Environ. Res. Risk Assess.* 21 (3), 287–297.
- Yu, L., Zhong, S., Pei, L., Bian, X., Heilman, W.E., 2016. Contribution of large-scale circulation anomalies to changes in extreme precipitation frequency in the United States. *Environ. Res. Lett.* 11, 044003.